

## Generatori di Variabili Casuali

Michela Meo  
Maurizio M. Munafò

Michela.Meo@polito.it - Maurizio.Munafò@polito.it

## Obiettivo

- I simulatori necessitano di generatori di istanze di variabili casuali con una data distribuzione
- Procediamo come segue:
  - Generiamo una sequenza  $X$  di numeri casuali interi
  - Partendo da  $X$ , generiamo una sequenza  $Z$  di istanze di variabile casuale uniforme in  $(0,1)$
  - Usando  $Z$ , generiamo istanze della v.c. di interesse

## Metodo della Trasformazione Inversa

- Si vogliono generare istanze di una variabile casuale  $X$  con distribuzione cumulativa  $F(x)$
- Il metodo della trasformazione inversa si basa sulla proprietà che  $u=F(x)$  è una v.c. uniformemente distribuita in  $[0,1]$
- Quindi si genera un'istanza  $u$  uniforme e poi si calcola

$$x=F^{-1}(u)$$

## Trasformazione Inversa

Dimostrazione che  $u=F(x)$  è uniforme in  $[0,1]$ :

- Sia  $Y=g(X)$  una funzione di  $X$  invertibile e monotona non decrescente,  $X=g^{-1}(Y)$ 

$$F_Y(y) = P(Y \leq y) = P(g^{-1}(Y) \leq g^{-1}(y))$$

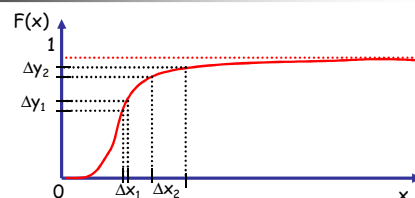
$$= P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$
- Scegliamo  $g(\cdot)$  tale che  $g(X)=F_X(X)$ , o anche  $Y=F_X(X)$ , e  $0 \leq y \leq 1$ ,
 
$$F_Y(y) = F_X(g^{-1}(y)) = F_X(F_X^{-1}(y)) = y$$

## Trasformazione Inversa

- Inoltre, da  $F_Y(y)=y$ ,
 
$$f_Y(y) = dF_Y / dy = 1$$
- quindi  $Y$  è uniforme in  $[0,1]$

**Generando  $Y$  uniforme in  $[0,1]$  e ricavandone l'inversa,  $F^{-1}(Y)$ , si ottiene  $X$**

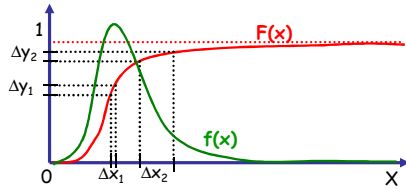
## Trasformazione Inversa



- Se  $\Delta y_1 = \Delta y_2$ , la probabilità di generare un'istanza di  $x$  in  $\Delta x_1$  o in  $\Delta x_2$  è la stessa
- A pari  $dx$ , i numeri in  $\Delta x_1$  sono più frequenti di quelli in  $\Delta x_2$ , perché  $\Delta x_1$  è più stretto

## Trasformazione Inversa

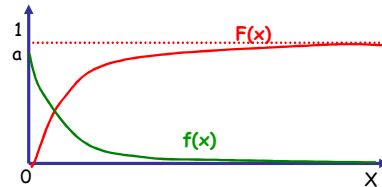
Gli intervalli sull'asse  $x$ , sono più stretti se la derivata di  $F(x)$  è grande, i.e.,  $f(x)$  è maggiore



## Esponenziale Negativa

- Si vogliono generare istanze di una variabile casuale esponenziale negativa a tasso  $a$ ,  $x \geq 0$ ,

- $f(x) = ae^{-ax}$ ,  $F(x) = 1 - e^{-ax}$



## Esponenziale Negativa

- L'inversa di  $F(x)$  è

$$y = 1 - e^{-ax}$$

$$e^{-ax} = 1 - y$$

$$-ax = \ln(1 - y)$$

$$x = -\frac{1}{a} \ln(1 - y)$$

## Esponenziale Negativa

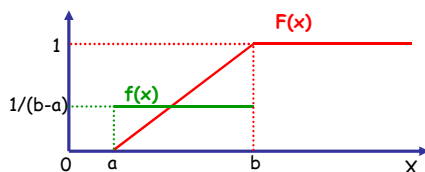
- Genera  $u = U(0,1)$ , uniforme in  $(0,1)$
- Calcola  $x = -1/a \ln(1-u)$
- Restituisci  $x$

- Poiché sia  $u$  che  $1-u$  sono uniformemente distribuiti in  $(0,1)$ ,  $x$  si può ricavare da  $x = -1/a \ln(u)$

## Uniforme

- Si vogliono generare istanze di una variabile casuale uniforme con supporto  $[a,b]$

- $f(x) = 1/(b-a)$ ,  $F(x) = (x-a)/(b-a)$ ,  $a \leq x \leq b$



## Uniforme

- L'inversa di  $F(x)$  è

$$y = \frac{(x-a)}{(b-a)}$$

$$(b-a)y = x - a$$

$$x = a + (b-a)y$$

- Genera  $u = U(0,1)$ , uniforme in  $(0,1)$
- Calcola  $x = a + (b-a)u$
- Restituisci  $x$

## Pareto

- Distribuzione Pareto
  - $f(x) = ax^{-(a+1)}$ ,  $F(x) = 1-x^{-a}$
- Inversa:

$$y = 1 - x^{-a}$$
$$x^{-a} = 1 - y$$
$$x = \frac{1}{(1-y)^{1/a}}$$

## Pareto

1. Genera  $u=U(0,1)$ , uniforme in  $(0,1)$
2. Calcola  $x = 1/u^{1/a}$
3. Restituisci  $x$

## Weibull

- Distribuzione Weibull
  - $F(x) = 1-e^{-(x/a)^b}$
- Inversa:

$$y = 1 - e^{-(x/a)^b}$$
$$e^{-(x/a)^b} = 1 - y$$
$$-\left(\frac{x}{a}\right)^b = \ln(1 - y)$$
$$x = a[-\ln(1 - y)]^{1/b}$$

## Weibull

1. Genera  $u=U(0,1)$ , uniforme in  $(0,1)$
2. Calcola  $x = a(-\ln(u))^{1/b}$
3. Restituisci  $x$

## Bernoulli

- Distribuzione Bernoulli
  - $P(X=x_1)=p$ ,  $P(X=x_2)=1-p$  e  $0 < p < 1$
  - Rappresenta eventi con due possibili valori (successo/insuccesso, vero/falso)
- Generazione
  1. Genera  $u=U(0,1)$ , uniforme in  $(0,1)$
  2. Se  $u \leq p$  restituisci  $x=x_1$
  3. altrimenti restituisci  $x=x_2$

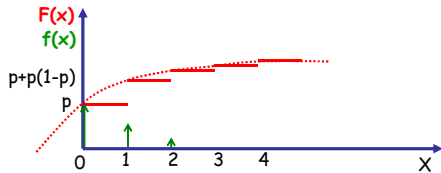
## Geometrica

- Distribuzione geometrica
  - La probabilità di  $x$  è  $P(X=x)=p(1-p)^x$ , con  $x=0,1,2,\dots$  e  $0 < p < 1$
  - La cumulativa è:  $P(X \leq x) = 1-(1-p)^{x+1}$
- Inversa:

$$y = 1 - (1-p)^{x+1}$$
$$(1-p)^{x+1} = 1 - y$$
$$(x+1)\ln(1-p) = \ln(1-y)$$
$$x = \frac{\ln(1-y)}{\ln(1-p)} - 1$$

## Geometrica

$$x = \frac{\ln(1-y)}{\ln(1-p)} - 1 \Rightarrow x = \left\lceil \frac{\ln(u)}{\ln(1-p)} - 1 \right\rceil$$



## Geometrica

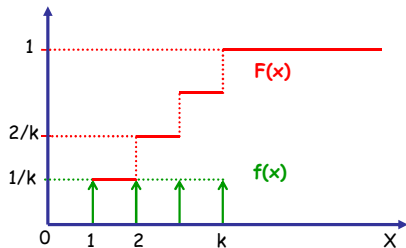
1. Genera  $u=U(0,1)$ , uniforme in  $(0,1)$
2. Calcola  $x = \text{ceil}(\ln(u)/\ln(1-p) - 1)$
3. Restituisci  $x$

- Per la geometrica con probabilità di  $x$   
 $P(X=x)=p(1-p)^{x-q}$ , con  $x=q,q+1,q+2,\dots$

$$x = q + \left\lceil \frac{\ln(u)}{\ln(1-p)} - 1 \right\rceil$$

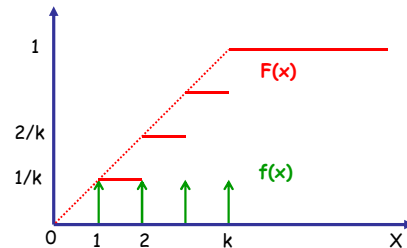
## Distribuzione Uniforme Discreta

- Si vuole generare un'istanza  $x$  di  $X$ , con  $x$  in  $\{1,2,\dots,k\}$  e  $p(X=x)=1/k$



## Distribuzione Uniforme Discreta

- Si ha  $F(x)=x/k$  con  $x=1,2,\dots,k$



## Distribuzione Uniforme Discreta

- L'inversa è  $x=yk$
- $X$  si può ottenere da  $x = \lceil ku \rceil$
- Infatti, avendo generato  $u$  uniforme in  $(0,1)$ , si restituisce  $x$  se

$$\frac{x-1}{k} < u \leq \frac{x}{k}$$

ossia

$$x-1 < ku \leq x$$

$$ku \leq x < ku+1$$

## Distribuzioni Empiriche o Discrete

- Si vogliono generare istanze di una variabile casuale la cui distribuzione è empiricamente calcolata tramite
  - Misure
  - Approssimazioni (non ammette l'inversa in forma chiusa)
- $X$  assume i valori  $x_1, x_2, \dots, x_k$  con probabilità  $p_1, p_2, \dots, p_k$
- Si ricava la cumulativa
- Si ricava l'inversa

## Distribuzioni Empiriche

- La distribuzione cumulativa si ricava empiricamente dai valori misurati o approssimati

Valori di $x$	$x_1$	$x_2$	$x_3$	...	$x_k$
Probabilità	$p_1$	$p_2$	$p_3$	...	$p_k$
CDF, $F(x)$	0 $0 \leq x < x_1$	$p_1$ $x_1 \leq x < x_2$	$p_1 + p_2$ $x_2 \leq x < x_3$	...	1 $x_k \leq x$

## Distribuzioni Empiriche

- Per generare istanze di  $X$ , si opera così:

- Genera  $u=U(0,1)$ , uniforme in  $(0,1)$
- Restituisci  $x$ , secondo la condizione soddisfatta da  $u$  nella tabella seguente

Valori di $x$	$x_1$	$x_2$	...	$x_k$
Condizione	$0 < u \leq F(x_1)$	$F(x_1) < u \leq F(x_2)$	...	$F(x_{k-1}) < u \leq F(x_k)$

## Metodo della Trasformazione Inversa

- Per essere applicabile si deve poter invertire la funzione  $F(x)$
- Richiede la generazione di un unico numero casuale  $U(0,1)$
- Il costo computazionale dipende dalla complessità di calcolo della funzione inversa

## Metodo della Convoluzione

- Si usa quando una variabile casuale si può esprimere come somma di altre variabili che siano facilmente generabili
 
$$x = y_1 + y_2 + \dots + y_n$$
- Si genera un'istanza per ogni  $y_i$
- Sommando le  $y_i$  si ottiene un'istanza di  $x$

## Metodo della Convoluzione

- Casi in cui si applica il metodo della convoluzione:
  - Erlang-K: è la somma di  $K$  variabili casuali con distribuzione esponenziale
  - Binomiale con parametri  $n$  e  $p$ : è la somma di  $n$  variabili Bernoulli con probabilità di successo  $p$
  - Chi-square con  $n$  gradi di libertà: è la somma del quadrato di  $n$  normali  $N(0,1)$

## Distribuzione Erlang-K

- La Erlang-K a media  $1/m$  è la somma di  $K$  variabili casuali con distribuzione esponenziale a media  $1/(Km)$ ,  $X = \sum_{i=1}^K X_i$
- Un'istanza  $x$  di  $X$  si ricava dalla somma di istanze  $x_i$  di  $X_i$

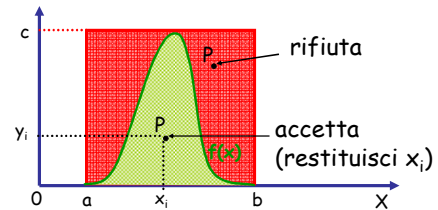
$$x = \sum_{i=1}^K \left( -\frac{1}{Km} \ln(u_i) \right) = -\frac{1}{Km} \ln \left( \prod_{i=1}^K u_i \right)$$

### Metodo Rifiuto/Acettazione

- Il metodo del rifiuto si applica a quelle variabili casuali con ddp  $f(x)$  con supporto limitato in  $[a,b]$
- Se  $c$  è il massimo valore assunto da  $f(x)$ , si opera come segue:
  - Genera  $x_i=U(a,b)$ , uniforme in  $[a,b]$
  - Genera  $y_i=U(0,c)$ , uniforme in  $[0,c]$
  - Se  $y_i \leq f(x_i)$  restituisci  $x_i$ , altrimenti torna al passo 1

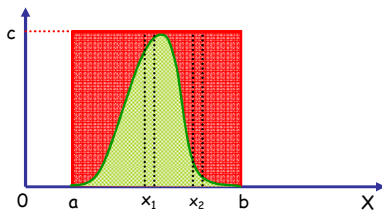
### Metodo Rifiuto/Acettazione

- Generare  $x_i=U(a,b)$  e  $y_i=U(0,c)$  significa generare casualmente un punto  $P$  nel rettangolo delimitato da  $(a,b)$  sull'asse  $x$  e  $(0,c)$  su  $y$



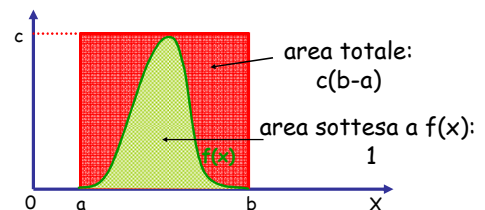
### Metodo Rifiuto/Acettazione

- La probabilità di accettazione in un intervallo  $(x_1, x_1+dx)$  dell'asse  $x$  è proporzionale a  $f(x_1)$



### Metodo Rifiuto/Acettazione

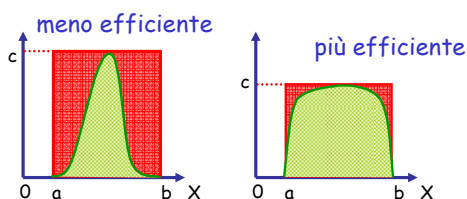
- L'efficienza del metodo dipende dall'area del rettangolo che racchiude  $f(x)$ , ossia da quanto il rettangolo è una buona approssimazione di  $f(x)$



### Metodo Rifiuto/Acettazione

- La probabilità  $P$  di accettazione di un punto generato a caso è pari al rapporto area sottesa a  $f(x)$ /area totale:

$$P = 1/[c(b-a)]$$



### Metodo Rifiuto/Acettazione

- Il numero medio di punti generati per ottenere un'istanza di  $X$  è pari a

$$N = \sum_{i=1}^{\infty} i(1-P)^{i-1}P = \frac{1}{P}$$

- Il metodo si può applicare a variabili a supporto infinito solo in modo approssimato, troncandone il supporto

## Metodo Rifiuto/Acettazione

- In generale, se esiste una ddp  $g(x)$  tale che  $kg(x) \geq f(x)$ , posso ottenere campioni di  $f(x)$  operando come segue:
  1. Genera  $x_i$  dalla distribuzione  $g(x)$
  2. Genera  $y_i = U(0, kg(x_i))$ , uniforme in  $[0, kg(x_i)]$
  3. Se  $y_i \leq f(x_i)$  restituisci  $x_i$ , altrimenti torna al passo 1
- La probabilità  $P$  di accettazione in questa forma generale è quindi  $P=1/k$

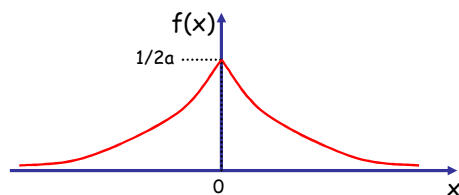
## Metodo di Composizione

- Il metodo si applica alle variabili casuali la cui CDF si può esprimere come somma pesata di altre funzioni CDF
$$F(x) = \sum_{i=1}^n p_i F_i(x), \quad \text{con } p_i > 0, \sum_{i=1}^n p_i = 1$$
- Genera un'istanza della v.c.  $I$ , tale che  $P(I=i) = p_i$
- Genera un'istanza di v.c. con CDF  $F_i(x)$
- Il metodo si può applicare anche alle ddp  $f(x)$

## Distribuzione di Laplace

- Si vuole generare  $X$  con ddp

$$f(x) = \frac{1}{2a} e^{-|x|/a}$$



## Distribuzione di Laplace

- La distribuzione di Laplace è una composizione di due esponenziali, con probabilità  $\frac{1}{2}$  è positiva e con la stessa probabilità è negativa
  1. Genera  $u_1 = U(0,1)$ , e  $u_2 = U(0,1)$
  2. Se  $u_1 < 1/2$  restituisci  $x = a \ln(u_2)$  altrimenti, restituisci  $x = -a \ln(u_2)$
- NB: questa distribuzione si può generare con la trasformazione inversa (è più efficiente)

## Altri Metodi

- Esistono diversi metodi specifici per alcune particolari distribuzioni
- Di solito questi metodi ad hoc sono basati su proprietà matematiche soddisfatte da certe variabili

## Distribuzione Normale

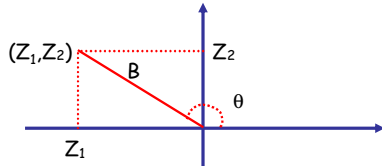
- Si vuole generare  $X$  con distribuzione normale con media 0 e varianza 1,  $N(0,1)$ ,

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

l'inversa non può essere scritta in forma chiusa

## Distribuzione Normale

- Si usano alcune proprietà della distribuzione normale
- Si considerino 2 v.c. normali viste come coordinate cartesiane di un punto nel piano



## Distribuzione Normale

- Esprimendo il punto in coordinate polari si ha
  - $Z_1 = B \cos \theta$
  - $Z_2 = B \sin \theta$
  - $B^2 = Z_1^2 + Z_2^2$
- Si generano separatamente  $B$  e  $\theta$  e si ricavano poi  $Z_1$  e  $Z_2$

## Distribuzione Normale

- La variabile  $B^2$  ha distribuzione chi-quadro con 2 gradi di libertà, ossia è esponenziale con media 2
- $B^2$  si genera con la trasformata inversa
  1. Genera  $u=U(0,1)$
  2. Calcola  $B= \sqrt{-2\ln(u)}$
  3. Restituisci  $B$

## Distribuzione Normale

- L'angolo  $\theta$  ha distribuzione uniforme in  $(0,2\pi)$  ed è indipendente da  $B$
- $\theta$  si genera a sua volta con la trasformata inversa
  1. Genera  $u=U(0,1)$
  2. Calcola  $\theta= 2\pi u$
  3. Restituisci  $\theta$

## Distribuzione Normale

La procedura è quindi la seguente

1. Genera  $u=U(0,1)$
2. Genera  $v=U(0,1)$
3. Calcola  $B=\sqrt{-2\ln(u)}$
4. Calcola  $\theta = 2\pi v$
5. Calcola  $Z_1 = B \cos(\theta)$
6. Calcola  $Z_2 = B \sin(\theta)$

## Distribuzione Normale

- La procedura genera coppie di istanze di una normale al costo della generazione di 2 istanze di v.c. uniforme e alcune operazioni algebriche
- Data l'istanza  $Z$  di v.c. normale  $N(0,1)$ , istanze di una v.c.  $X$  normale con media  $\mu$  e varianza  $\sigma^2$  si ricavano da

$$X = \mu + \sigma Z$$

## Distribuzione Poisson

- Si vogliono generare istanze di una v.c.  $X$  con distribuzione Poisson con media  $a$ ,

$$p(n) = P(X = n) = \frac{e^{-a} a^n}{n!}, \quad n = 0, 1, 2, \dots$$

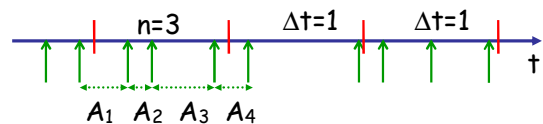
- $X$  rappresenta il numero di arrivi di un processo di Poisson in un'unità di tempo, se  $a$  è il numero medio di arrivi nell'u.t.

## Distribuzione Poisson

- $X=n$  se e solo se vale la relazione

$$A_1 + A_2 + \dots + A_n \leq 1 < A_1 + A_2 + \dots + A_n + A_{n+1}$$

dove  $A_i$  è l' $i$ -esimo tempo di interarrivo



## Distribuzione Poisson

- I tempi di interarrivo  $A_i$  hanno distribuzione esponenziale negativa con media  $1/a$
- Si può quindi generare un'istanza di  $X$  generando i tempi di interarrivo e calcolando quanti arrivi si verificano nell'unità di tempo

## Distribuzione Poisson

$$A_1 + \dots + A_n \leq 1 < A_1 + \dots + A_n + A_{n+1}$$

$$\sum_{i=1}^n -\frac{1}{a} \ln(u_i) \leq 1 < \sum_{i=1}^{n+1} -\frac{1}{a} \ln(u_i)$$

$$\sum_{i=1}^n \ln(u_i) \geq -a > \sum_{i=1}^{n+1} \ln(u_i)$$

$$\ln\left(\prod_{i=1}^n u_i\right) \geq -a > \ln\left(\prod_{i=1}^{n+1} u_i\right)$$

$$\prod_{i=1}^n u_i \geq e^{-a} > \prod_{i=1}^{n+1} u_i$$

## Distribuzione Poisson

La procedura è quindi la seguente

- Inizializza:  $n=0$ ,  $q=1$
- Genera  $u_{n+1}=U(0,1)$
- $q=qu_{n+1}$
- Se  $q < e^{-a}$   
restituisce  $n$ ,  
altrimenti  
 $n=n+1$  e torna al passo 2

## Distribuzione Poisson

- Il numero medio di istanze di uniforme necessarie per la generazione di un'istanza di  $X$  è dato da

$$\begin{aligned} N &= \sum_{n=0}^{\infty} (n+1)p(n) = \sum_{n=0}^{\infty} (n+1) \frac{e^{-a} a^n}{n!} = \\ &= \sum_{n=0}^{\infty} n \frac{e^{-a} a^n}{n!} + \sum_{n=0}^{\infty} \frac{e^{-a} a^n}{n!} = \sum_{n=1}^{\infty} \frac{e^{-a} a^n}{(n-1)!} + 1 = \\ &= a \sum_{n=0}^{\infty} \frac{e^{-a} a^n}{n!} + 1 = a + 1 \end{aligned}$$

## Distribuzione Binomiale

- Si vogliono generare istanze di una v.c.  $X$  con distribuzione binomiale  $(p,n)$

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n$$

- $X$  rappresenta il numero di successi in  $n$  esperimenti indipendenti con probabilità di successo  $p$

## Distribuzione Binomiale

### Metodo di generazione 1 (convoluzione).

- Genera  $n$  istanze  $u_i$  di  $U(0,1)$
- Conta il numero  $x$  di variabili  $u_i$  che sono minori di  $p$
- Restituisci  $x$

## Distribuzione Binomiale

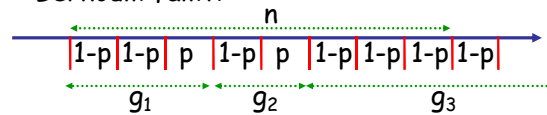
### Metodo di generazione 2 (inversa).

- Calcola  $F(x)$  e memorizzala nell'array  $A$
- Genera  $u=U(0,1)$
- Cerca  $x$  tale che  $A[x] \leq u < A[x+1]$
- Restituisci  $x$

## Distribuzione Binomiale

### Metodo di generazione 3.

- Si basa sull'osservazione che una geometrica a parametro  $(1-p)$  che assume il valore  $i+1$  corrisponde a  $i$  esperimenti Bernoulli falliti



$$g_1 + g_2 + g_3 > n \rightarrow \text{restituisci } 2$$

## Distribuzione Binomiale

### Metodo di generazione 3.

- Inizializza:  $m=1, q=0$
- Genera  $u_m=U(0,1)$
- Genera geometrica  $g_m = \text{ceil}(\ln(u_m)/\ln(1-p))$
- Calcola  $q=q+g_m$
- Se  $q > n$   
restituisci  $m-1$ ,  
altrimenti  
 $m=m+1$  e torna al passo 2